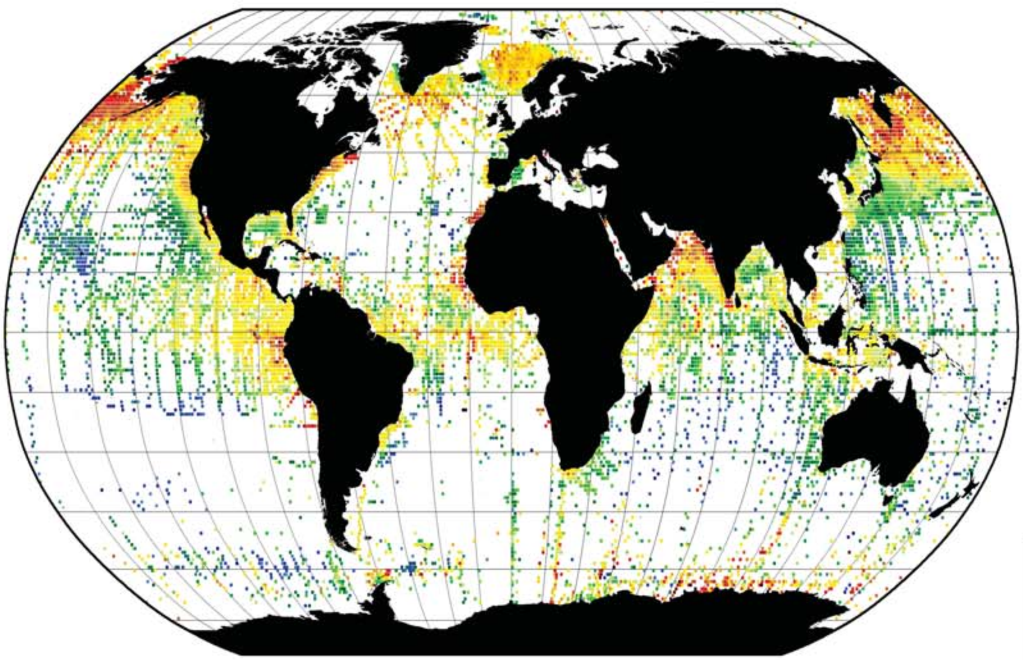
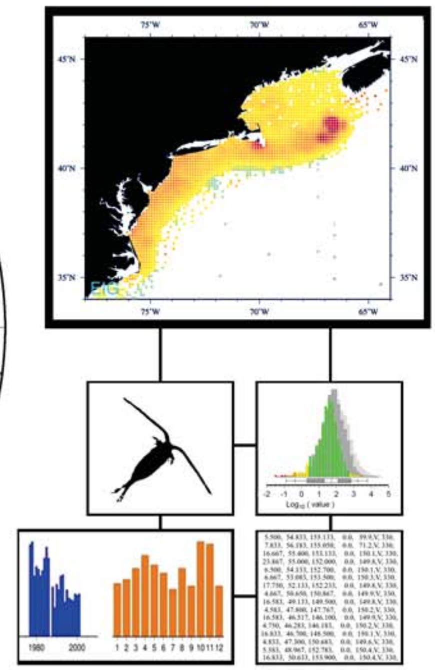
NMFS-COPEPOD

The **C**oastal & **O**ceanic **P**lankton **E**cology, **P**roduction & **O**bservation **D**atabase





COPEPOD: The Global Plankton Database

An overview of the 2010 database contents, processing methods, and access interface.

2010

Office of Science & Technology
National Marine Fisheries Service
National Oceanic and Atmospheric Administration
U.S. Department of Commerce

NOAA Technical Memorandum NMFS-F/ST-36
December 2010

About this document:

The mission of the National Oceanic and Atmospheric Administration (NOAA) is to understand and predict changes in the Earth's environment and to conserve and manage coastal and oceanic marine resources and habitats to help meet our Nation's economic, social, and environmental needs. As a branch of NOAA, the National Marine Fisheries Service (NMFS) conducts or sponsors research and monitoring programs to improve the scientific basis for conservation and management decisions. NMFS strives to make information about the purpose, methods, and results of its scientific studies widely available.

The NMFS Office of Science & Technology (NMFS-ST) uses the **NOAA Technical Memorandum** series to achieve timely dissemination of scientific and technical information that is of high quality but inappropriate for publication in the formal peer-reviewed literature. The contents are of broad scope, including technical workshop proceedings, large data compilations, status reports and reviews, lengthy scientific or statistical monographs, and more. NOAA Technical Memoranda published by the NMFS-ST, although informal, are subject to extensive review and editing and reflect sound professional work. Accordingly, they may be referenced in the formal scientific and technical literature.

This document should be cited as follows:

O'Brien, T.D. 2010. COPEPOD: The Global Plankton Database. An overview of the 2010 database contents, processing methods, and access interface. U.S. Dep. Commerce, NOAA Tech. Memo., NMFS-F/ST-36, 28 pp.

For further information direct inquiries to

Todd D. O'Brien
NOAA – NMFS – F/ST7
1315 East-West Highway - #12535
Silver Spring, Maryland 20910

Email: Todd.OBrien@noaa.gov

Phone: 301-427-8160

This document and the NMFS-COPEPOD database and data products are available online at:

http://www.st.nmfs.noaa.gov/plankton

COPEPOD: The Global Plankton Database

An overview of the 2010 database contents, processing methods, and access interface.

Todd D. O'Brien

NOAA Technical Memorandum NMFS-F/ST-36
December 2010

U.S. Department of Commerce
Gary Locke, Secretary

National Oceanic and Atmospheric Administration
Jane Lubchenco
Under Secretary for Oceans and Atmosphere

National Marine Fisheries Service
Eric C. Schwaab, Assistant Administrator for Fisheries

COPEPOD

The global plankton database!

COPEPOD: The Global Plankton Database

Todd D. O'Brien
NOAA - National Marine Fisheries Service (NMFS)
Office of Science & Technology - Marine Ecosystems Division
Silver Spring, Maryland

The **C**oastal & **O**ceanic **P**lankton **E**cology, **P**roduction & **O**bservation **D**atabase (**COPEPOD**) is an online, global-coverage database of zooplankton and phytoplankton abundance, biomass, and composition data. COPEPOD provides “raw” plankton data and integrated data products through its plankton-tailored online interface. Through a variety of access levels and packaging, COPEPOD preserves the individual identities of its data pieces while also offering regional and global compilations. While new COPEPOD data are made available online each month, this document summarizes the content of the database as of December 2010. Hereafter known as *COPEPOD-2010*, the processing methods presented in this document reflect the current configuration and layout of the active COPEPOD data system. This configuration replaces all earlier versions of the database (2007, 2005) and will remain the active version until the next update is released in December of 2012.

1. INTRODUCTION

In 2004, a new plankton database effort called “COPEPOD” was initiated to consolidate and integrate the sixty years of plankton monitoring done by the National Marine Fisheries Service (and its earlier institutional entities and forms) with data from other fisheries and research institutes from around the world. COPEPOD is a continuously updated and growing data management effort, incorporating over fourteen years of plankton data management experience into designing and building an online data system designed specifically for serving plankton data to the plankton and oceanographic scientists that use these data. COPEPOD focuses on preserving the identity of the individual data sets, highlighting each with a full summary of their exact content, sampling methods, and investigators associated with those data. By further packaging these individual sets into larger data compilations and data products, a user can work at variety of local, regional, or global scales. Whether looking for data from a specific project or across an entire basin, COPEPOD serves quality data via a user friendly interface and in a variety of data formats.

While *COPEPOD-2010* includes raw content originally introduced in earlier versions of the database (*e.g.*, *COPEPOD-2007, COPEPOD-2005*), there have been changes to the supplemental indexing (e.g., the Plankton Group Code, Section 3.2) and the quality control methods (Section 4, Table 5). The changes have been applied to all existing data in the COPEPOD database, per the new *COPEPOD-2010* methodologies summarized in this document. New data and updates from ongoing NMFS monitoring programs have also been added to existing COPEPOD data collections (*e.g.*, CalCOFI, EcoFOCI, EcoMon, SEAMAP). Users of earlier versions of the COPEPOD database are therefore encouraged to consider *COPEPOD-2010* as a replacement and should update from these earlier database releases.

2. DATA SOURCES

COPEPOD-2010 content represents the ongoing efforts of over fourteen years of plankton data discovery, acquisition, compilation, and processing. This content comes from ongoing NMFS Ecosystem Surveys (and other NMFS associated sampling programs), from data rescued by the Historical Plankton Data Search & Rescue project (*COPEPOD-SAR*), from various institutional & project-funded sampling programs, and from investigators who have submitted data from their work directly to COPEPOD for inclusion in the effort (Figure 1).

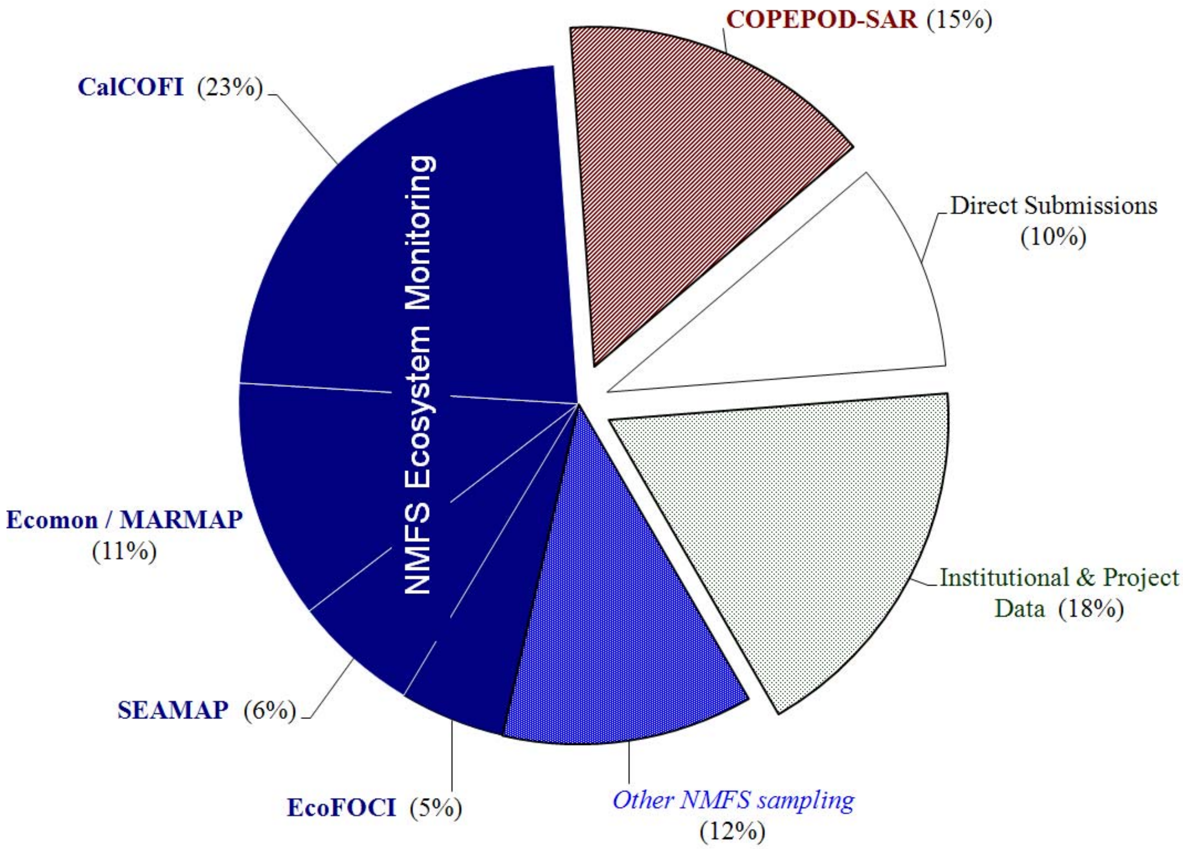


Figure 1: Plankton data sources, shown as the percentage of all tows[1] in the COPEPOD database as of December 2010. NMFS associated sampling represents 57% of the global tow data.

[1] Traditionally, data presence in COPEPOD has been represented using the number of net tows present in each data set. A "net tow" may contain a single total biomass value, or it may contain multiple species abundance values. An alternate way to represent each data set is in terms of the number of plankton observations (measurements) recorded. COPEPOD's total plankton observation content is dominated by just four data sets which represent over 60% of the global observations: IIOE (*International Indian Ocean* Expedition), PETAO (*Pelagic Ecosystems of the Atlantic Ocean*), CSK (*Cooperative Study of the Kuroshio and adjacent regions*), and CalCOFI (*California Cooperative Oceani Fisheries Investigations*). Representing IIOE in terms of tows, the data set contains less than 2,000 net tows. As each of these net tows contains up to 100 taxa observations/measurements, IIOE dominates COPEPOD with a total of over 180,000 observations.

2.1 The NMFS Ecosystem Surveys

The National Marine Fisheries Service (NMFS) regularly samples plankton as part of its ongoing Ecosystem Survey programs (*Figure 2*). These regional sampling efforts include zooplankton displacement volumes as well as zooplankton and ichthyoplankton composition and abundance samples.

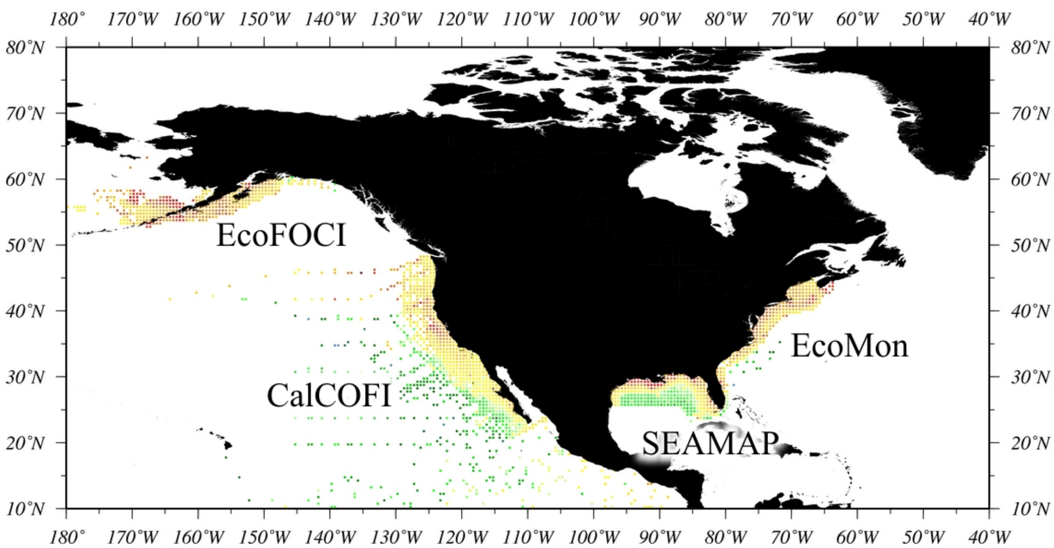


Figure 2: Map of zooplankton biomass and associated programs of the NMFS Ecosystem Surveys

CalCOFI	- **Cal**ifornia **C**ooperative **O**ceanic **F**isheries **I**nvestigations (*1951- present*)
EcoFOCI	- **Eco**systems & **F**isheries-**O**ceanography **C**oordinated **I**nvestigations (*1979-present*)
EcoMon	- **Eco**system **Mon**itoring - formerly known as MARMAP (*1977-present*)
SEAMAP	- **S**outh**Ea**st **M**onitoring & **A**ssessment **P**rogram (*1982-present*)

NMFS also has a rich history of zooplankton sampling going back to the early 1950's. Known in early years as the Bureau of Commercial Fisheries (BCF), NMFS/BCF has repeatedly worked together with State and local research institutions to sample coastal US waters and Hawaii, and regularly participated in large projects such as EASTROPAC which sampled far out into the equatorial Pacific (Figure 3).

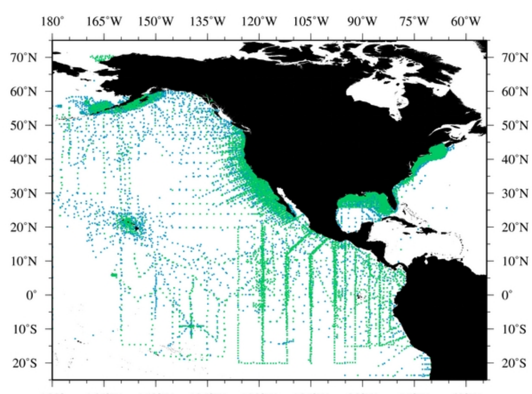


Figure 3: Map of historical and ongoing NMFS-associated plankton-sampling programs and surveys.

2.2 Historical Plankton Data Search & Rescue

Large amounts of historical plankton data still exist only in non-electronic, paper formats. Some of these documents are available through journals or in library collections, but larger amounts remain inaccessible, hidden in institutional archives, investigators' filing cabinets and storage rooms.

Over the last ten years, ongoing digitization efforts have managed to key large amounts of these data into an electronic form. In addition to the COPEPOD author's ten years of effort, other institutional and personal efforts have made large collections of historical Japanese (*e.g.*,Odate time series, Hokkaido University surveys) and Former Soviet Union (*e.g.*, Sergey Piontkovski's *Global Plankton Database* project) plankton data readily available to the scientific community.

The COPEPOD *Historical Plankton Data Search & Rescue* projec*t* (COPEPOD-SAR) is actively keying paper plankton data sources into a digital form, quality checking them, and then distributing them online via the COPEPOD database. This past and ongoing keying work has been accomplished through funding from NOAA's Climate Data Modernization Program (CDMP) and through NOAA's Environmental Services Data and Information Management (ESDIM) program (Table 1).

Data Rescue Funding Source	**Project Duration**	**Principal Investigator(s)**
CDMP	5 years (2005 – present)	*T.D. O'Brien*
ESDIM	2 years (2001 – 2002) 2 years (2003 – 2004)	*T.D. O'Brien*
ESDIM	2 years (1999 – 2000)	*T.D. O'Brien* *& M.E. Conkright*

Table 1: Summary of COPEPOD plankton data search and rescue funding.

2.3 Data Centers, Institutions, & Project Data

Globally sampled plankton data were acquired in various electronic formats from international and regional data centers (*e.g.,* the British Oceanographic Data Centre, the Indian National Oceanographic Data Center, the Japanese Oceanographic Data Center, the U.S. National Oceanographic Data Center), research institutions (*e.g.,* Woods Hole Oceanographic Institute, Smithsonian Institution, Sir Alister Hardy Foundation for Ocean Science), and various oceanographic projects (*e.g.,* JGOFS, GLOBEC).

2.4 Direct Investigator Submission

An increasing amount of plankton data content is now being offered directly to COPEPOD by the original collecting investigators. We especially welcome these data as the direct correspondence with the collecting investigator(s) allows us to insure we correctly represent the methods and data, and that we properly acknowledge the responsible investigators and institutions. As a result, these submissions tend to be the most recent and highest quality data in the COPEPOD database.

Whether it is a single cruise or a collection, COPEPOD invites more investigators to consider sharing their data through COPEPOD. Our goal is to provide quality data in an easy-to-access, easy-to-use database that also retains credit and acknowledgement for those whose hard work collected the data.

Appendix I, located at the back of this document, lists all *known* investigators whose work is present in COPEPOD as of December 2010. Unfortunately, information on the collecting investigators or institutions is not always known or available. When possible we try to track down this information and add it to the database and content summaries, but this is not always possible. COPEPOD welcomes feedback from the user community regarding missing or incomplete metadata, and will gladly make correction or additions when this information is provided.

3. DATA PROCESSING

The hundreds of plankton data sets present in COPEPOD were originally stored in hundreds of different files, formats, and data structures (*e.g.*, tables, spreadsheets, ASCII text files). The first and most time-consuming step in adding these data to COPEPOD involves carefully translating these formats and variables into the COPEPOD variable definition set and data structure. This step also involves reviewing the original data documentation to ensure that all of the methods and metadata are accurately represented during the translation and loading. Once these steps are completed, the original data values are available in the standard COPEPOD electronic format, but they do not always have comparable measurement units or a standard taxonomic indexing system. The second major step in loading these data into COPEPOD is therefore to verify and classify all of the plankton taxa into general taxonomic groupings (Sections 3.1-3.2) and then calculate (where needed) common base-unit values from all of the original data values (Section 3.3).

3.1 Taxonomic Translation and Name Verification

COPEPOD's approach to taxonomic name management is to preserve the original, investigator-provided, description as completely and correctly as possible. The spelling and taxonomic validity of each taxonomic name is checked against the Integrated Taxonomic Information System (ITIS, *http://www.itis.gov*) and stored in the COPEPOD database using both the ITIS-generated unique taxonomic serial number (TSN) and the ITIS-adopted spelling for that taxonomic entity. In cases of taxonomic synonyms, the original description is retained and the record is stored with a second TSN that represents the currently accepted/valid name. This allows a COPEPOD user to use or ignore any taxonomic synonym corrections suggested by ITIS for a given taxonomic entity. If a taxonomic name is not found in ITIS, it is assigned a temporary identifier (from a sequence of unique, negative numbers maintained by COPEPOD) until it is reviewed by and added to the ITIS database. Finally, a separate sequence of unique, negative identifiers is used to represent non-taxonomic or multi-taxonomic descriptions (*e.g.*, "salps & doliods", "jellyfish", "*Calanus helgolandicus / Calanus carinatus*").

Non-taxonomic descriptive identifiers and/or modifiers are also frequently present in the original plankton observation description. These additional descriptors may include life stage information (*e.g.*, "adult", "juvenile", "copepodite C3"), gender (*e.g.*, "*male*", "*female*"), size ranges (*e.g.*, "> 1 mm", "2 – 5 µm", "< 40 µm"), and/or various taxonomic ranking indicators (*e.g.*, "spp.", "other", "miscellaneous"). These additional descriptors are also in the COPEPOD data structure to further preserve the original observation description.

3.2 Plankton Group Assignment

The COPEPOD database contains over 1.5 million plankton observations with over 5,000 unique taxonomic identifiers that range from individual species names (*e.g.*, *Calanus finmarchicus*, *Nitzschia delicatissima*) to general classes and families (*e.g.*, "calanoid copepods", "pennate diatoms"). To allow for quick and simple access to all data from a general plankton grouping such as "diatoms" or "copepods" or even something as broad as "all phytoplankton", a *Plankton Grouping Code* "smart-index" is added to every taxonomic record in the COPEPOD database. This smart-index is stored in addition to the original taxonomic descriptions and/or modifiers already described in Section 3.1, and greatly speeds up searching for and/or compiling data from entire groups of plankton using a simple numeric code.

The *Plankton Grouping Code* (PGC) is a seven digit number which identifies the plankton taxa's membership in up to four groups (see Table 2). For example *Calanus finmarchicus*, a calanoid copepod, is assigned a PGC of "4212010". This PGC indicates that *Calanus finmarchicus* is a "zooplankton" (Major Group = 4, in [4]212010), a "crustacean" (Minor group = 21, in 4[21]2010), a "copepod" (Focus Group = 20, in 421[20]10), and a "calanoid copepod" (Special Group = 10, in 42120[10]). Each and every database record for *Calanus finmarchicus* will have this PGC code (4212010), as will any other "calanoid copepod" species in the database. A euphausiid species has a PGC of 4218000, indicating it is a zooplankton (Major Group 4) and a crustacean (Minor Group 21) but that belongs to the euphausiid focus group (Focus Group 80) instead of the copepod focus group (Focus Group 20).

A single PGC value can be used to quickly identify and access an entire taxa group that contains hundreds of different species. Without this added code, a user would have to search millions of records and then sort through thousands of taxonomic names to identify and determine if it belonged to the specific group desired.

PGC	**Major Group** (##○○○○○○○)	**Minor Group** (○○##○○○○)	**Focus Group** (○○○○##○○)	**Special Group** (○○○○○○○##)	**Scientific Name**
1050000	Bacterioplankton	Cyanobacteria	-	-	*Oscillatoria thiebautii*
2160000	Phytoplankton	Diatom	-	-	*Skeletonema costatum*
4210**00**0	Zooplankton	Crustaceans	-	-	*Crustacea*
4210**00**0	Zooplankton	Crustaceans	Copepods	-	*Copepoda* spp.
4210**00**0	Zooplankton	Crustaceans	Copepods	Calanoid copepods	*Calanoida* spp.
4210**00**0	Zooplankton	Crustaceans	Copepods	Calanoid copepods	*Calanoides* spp.
42120**10**	Zooplankton	Crustaceans	Copepods	Calanoid copepods	*Calanus finmarchicus*
42120**10**	Zooplankton	Crustaceans	Copepods	Calanoid copepods	*Metridia pacifica*
42120**40**	Zooplankton	Crustaceans	Copepods	Cyclopoid copepods	*Oithona similes*
421**80**00	Zooplankton	Crustaceans	Euphausiids	-	*Euphausia pacifica*
4**32**0000	Zooplankton	Chaetognaths	-	-	*Parasagitta elegans*

Table 2: Examples of Plankton Grouping Codes (PGC) and inherent sub-grouping.

The four major, minor, and focus grouping levels of the PGC smart-index can be accessed by using simple *integer math*[2]. For example:

Major Groups can be selected by dividing any PGC by the 1000000:

If (PGC / 1000000) = 4 { it is a zooplankton }
If (PGC / 1000000) = 2 { it is a phytoplankton }

Example: "4212010" / 1000000 = 4 (*Calanus finmarchicus* is a zooplankton)

Minor groups can be selected by dividing any PGC by 10000:

If (PGC / 10000) = 216 { it is a diatom }
If (PGC / 10000) = 421 { it is a crustacean }
If (PGC / 10000) = 432 { it is a chaetognath }

Example: "4212010" / 10000 = 421 (*Calanus finmarchicus* is a crustacean)

Focus Groups can be selected by dividing by 100:

If (PGC / 100) = 42120 { it is a copepod }
If (PGC / 100) = 42180 { it is an euphausiid }

Example: "4212010" / 100 = 42120 (*Calanus finmarchicus* is a copepod)

The active Plankton Grouping Codes used in *COPEPOD-2010* are listed in Table 3.

[2] In integer math, the result of any calculation is rounded down to the next lowest integer, *regardless of the decimal value behind it*. Unlike the typical rounding you use in your daily life, in integer math a value of "1.6" or even "1.9999" will still round down to "1". This is a fundamental property behind mathematically applying and determining PGC group membership in software or programs outside of the COPEPOD web interface.

In many software and programming languages, enclosing the division calculation in "INT()" will force an integer-based calculation. For example: `4286000/10000 = 428.6` but `INT(4286000/10000) = 428`

PGC	MAJOR	Minor Groups	*Focus Groups*	*Special*
1000000	**Bacterioplankton**			
1050000		Cyanobacteria		
2000000	**Phytoplankton**			
2040000		Granuloreticulosa (foraminifera)		
2070000		Dinomastigota (dinoflagellates)		
2080000		Ciliophora (ciliates)		
2100000		Haptomonada (coccolithophores)		
2110000		Cryptomonada (cryptophytes)		
2120000		Discomitochondira (flagellates)		
2130000		Chrysomonada (chrysophytes)		
2160000		Diatoms		
2270000		Actinopoda (radiolarians)		
2280000		Chlorophyta (green algae)		
4000000	**Zooplankton**			
4030000		Cnidaria		
4032000			*Hydrozoa*	
4036000			*Scyphozoa*	
4038000			*Anthozoa*	
4040000		Ctenophora		
4050000		Platyhelminthes		
4090000		Nemertina		
4100000		Nematoda		
4130000		Rotifera		
4190000		Chelicerata		
4200000		Mandibulata		
4210000		Crustacea		
4211000			*Ostracoda*	
4212000			*Copepoda*	
4212010				*Calanoida*
4212030				*Harpacticoida*
4212040				*Cyclopoida*
4212050				*Monstriloida*
4212060				*Caligoida*
4213000			*Cirripedia*	
4000000	**Zooplankton** (*continued*)			
4214000			*Mysidacea*	
4216000			*Isopoda*	
4217000			*Amphipoda*	
4217025				*Gammaridea*
4217050				*Hyperiidea*
4217075				*Caprellidea*
4218000			*Euphausiacea*	
4219000			*Decapoda*	
4220000		Annelida		
4225000			*Polychaetes*	
4230000		Sipuncula		
4260000		Mollusca		
4262500			*Gastropoda*	
4265000			*Bivalvia*	
4266000			*Scaphopoda*	
4267500			*Cephalopoda*	
4290000		Bryozoa		
4300000		Brachiopoda		
4310000		Phoronida		
4320000		Chaetognatha		
4330000		Hemichordata		
4340000		Echinodermata		
4342000			*Asteroidea*	
4344000			*Ophuiroidea*	
4346000			*Echinodea*	
4348000			*Holothuroidea*	
4350000		Urochordata		
4352500			*Ascidiacea*	
4355000			*Thaliacea*	
4355010				*Salps*
4355050				*Doliolids*
4357500			*Appendicularia*	
4360000		Cephalochordata		
5000000	**Ichthyoplankton**			

Table 3: Active *Plankton Grouping Codes* (PGC) present in *COPEPOD-2010*. Groupings are based on Margulis and Schwartz (1998).

.

3.3 Calculation of Common Base-unit Values

The original plankton measurements in COPEPOD came in a variety of different measurement units (*e.g.*, "number per ml", "number per 30-liter sample", "number per 100 m^3", "number per m^2", "number per total net sample"). Within COPEPOD, these values are stored as they were originally measured with the following minor adjustments:

- Bulk multipliers in the unit numerator or denominator were calculated out to be just "number per unit" (*e.g.*, "35 critters **per 1000 m^3**" = "0.035 critters **per m^3**"; "3.5 **x 10^4** critters **per m^2**" = 35,000 critters **per m^2**");
- For microbial plankton (bacteria) data, base units (*e.g.*, "per ml", "per liter", "per m^3") were changed if necessary to keep the original values less than $1x10^9$ (*e.g.*, "bacteria = **1.5 x 10^9** per m^3" is stored as "bacteria = **1500** per ml").

Even after these simple adjustments above, the variety and types of original units still do not allow for instant inter-comparison of the data. One can not immediately compare data in units of "#/m^3" with data in units of "#/total-sample". To allow for easier use and intercomparison of different unit values like this, a *Common Base-unit Value* (CBV) was calculated and stored along with each original value. In COPEPOD-2010, this CBV is available in both "per-volume" (CBV-m3) and "per area" (CBV-m2) formats, with base units being assigned according to the plankton group and original measurement type present (Table 4).

Measurement Type & Group	CBV-m3 unit	CBV-m2 unit
Biomass (wet mass, dry mass, AFDM)	mg / m^3	mg / m^2
Biovolume (displacement volume, settled volume)	ml / m^3	ml / m^2
Zooplankton Abundance	# / m^3	# / m^2
Phytoplankton Abundance	# / l	# x 10^6 / m^2
Planktonic Bacteria (*Bacterioplankton*) Abundance	# / ml	# x 10^9 / m^2
Ichthyoplankton Abundance	# / m^3	# / m^2

Table 4: Measurement types and units for Common Base-unit "per volume" (CBV-m3) and "per area" (CBV-m2) Values.

If the original value was already in the correct CBV-m3 or CBV-m2 units, no calculation was necessary to create that CBV type. Otherwise, the common base unit value was calculated using the metadata associated with that sample. For example, "per total-net-sample" (per haul) measurements can be converted to CBV-m3 values by using flow meter 'volume of water filtered' data. If volume of water filtered was not provided with the original data, the "volume of water filtered" can often be estimated by multiplying the mouth area of the net opening by the distance the net was towed through the water column. For vertical tows, this towing distance is the lower depth minus the upper depth. For horizontal tows, towing distance are estimated by using the average towing speed and tow duration. A methods codes indicating the method used to calculate the CBV-m3, CBV-m2, or volume filtered value (if not provided) is stored in the database alongside each CBV value.

4. "Quality Control"

One of the biggest challenges in building a database of plankton sampled from a wide variety of sampling methods and sampling gear is how to quality check the wide variety data types and values found in this database. COPEPOD uses a numerical range checking system that quickly compares newly loaded plankton observation values with the thousands of other equivalent plankton observation values already present in the database. After this semi-automated review is completed, those new data can be immediately released and the new data values can be added to the existing range checking "data pool" to improve future ranging checks. The COPEPOD quality control data pool is therefore constantly improving as new data are added to the system. In addition to new data, new quality checks and methods are also regularly added (see Table 5).

The main purpose of COPEPOD's "quality control" check is to look for errors in the database translation and loading process. In general, plankton data were usually "correct" in their original source media and any anomalous values found in these original data are often due to natural processes (*e.g.*, blooms, swarms, patchiness) or mechanical sampling issues (*e.g.*, gear failure or clogged nets). The original authors often annotated these mechanical or bloom events within the original data documentation or data tables, but these annotations may not have been passed along when the data were distributed. The process of putting new plankton data into the database itself is typically the biggest reason for errors in the database, ranging from metadata mis-translation (*e.g.,* in a foreign-language document, did the author mean "*millimeters*" or "*micrometers*" by the first "m" in the units label "mm"), mislabeling of data types (*e.g.,* the data file say "per m^3", but the original methods documentation says "per 1000 m^3"), and a variety of numeric uncertainties (*e.g.,* "Is the comma in "1,234" a thousands indicator or a decimal indicator?"). In each of these examples, the value ranging question is not "Is this value 5.6 or 5.7?" but rather "Is this value 5.6 or 5600 or 0.0056?". These large differences are fairly easy to detect with automated ranging checks if the system is correctly comparing equivalent data types (*e.g.,* apples to apples[3], oranges to oranges).

4.1 Value Categorization and Sub-Grouping

The distribution and concentration of plankton in the water column varies by region, by season, and even by time of day. The sampling gear and methods used to collect the plankton also play a major role in exactly how much and what members of the plankton community are actually captured. Finally, how these samples are processed varies from investigator to investigator, with some investigators processing the sample to each individual species (*e.g.,* "100 *Calanus finmarchicus*", "50 *Acartia longiremis*", "50 *Calanus* other") and others only processing to broad groups (*e.g.,* "200 copepods"). Any attempt to set general ranges for these data values must therefore begin by taking into account factors such as the net mesh size, the season or month, the oceanographic region, and the taxonomic resolution and binning of the samples.

---

[3] The COPEPOD quality control system goes a step farther than "comparing apples to apples". Incoming plankton observation values are checked at a level equivalent to comparing "Washington-state Braeburn apples harvested in August" to other "Washington-state Braeburn apples harvested in August". This check can also be expanded to compare Washington-state Braeburn apples harvested in different months, or to compare different types of Washington-state apples, or even to compare Braeburn apples harvested in August in the United States to those harvested anywhere else in the world. In the end, the quality control system looks for claims of pumpkin-size apples being harvest in a region where most apples are typically cherry sized. The suspicious value is then flagged, but not removed, for further review.

COPEPOD currently contains over 2 million plankton biomass and abundance values, with observations from over 5,000 taxa, sampled by hundreds of different gear and mesh sizes in various regions around the world. While *COPEPOD-2005* and *WOD-2001* used very broad value classes for simple range checking, *COPEPOD-2010* introduces the next generation of plankton data range checking through advanced categorization and grouping (Table 5).

	WOD-2001 (O'Brien *et al.* 2002) ***WOD-2005**** ***WOD-2009****	**COPEPOD-2005** (O'Brien 2005)	**COPEPOD-2007** (O'Brien 2007)	**COPEPOD-2010**
Taxonomic Resolution	3 broad taxa groups + 5 biomass types	21 Major BGC Groups + 5 biomass types	85 MajorMinor PGC groups + 7 taxa-hierarchy levels + 4 life stage categories +6 biomass types	85 MajorMinor PGC groups + 7 taxa-hierarchy levels + 4 life stage categories +6 biomass types
Spatial Resolution	Global	Global	Global + 15 regions	Global + 15 regions
Temporal Resolution	"Annual"	"Annual" + 4 seasons	"Annual" + 4 seasons + Day/Night	"Annual" + 4 seasons + 12 months + Day/Night
Mesh Sizes	None	None	6 mesh size categories	6 mesh size categories

* WOD-2005 and WOD-2009 use the basic WOD-2001 quality control ranging approach, with adjusted numerical limits.

Table 5: History and complexity of plankton data quality control methods in WOD and COPEPOD.

COPEPOD-2010 Quality Control Sub-Categories:

The first step in the *COPEPOD-2010* advanced plankton quality control and value ranging system is to separate out each and every plankton observation value into a series of sub-categories that best represents the actual value type and sampling methods used for acquiring that value. The following sub-categories were used for ranging the *COPEPOD-2010* data:

Plankton Group: As mentioned in Section 3.2, each plankton observation is assigned to its corresponding plankton groups, such as "diatoms" or "copepods" or "chaetognaths", and assigned a *Plankton Grouping Code* (PGC) identifier. The *COPEPOD-2010* ranging system currently divides the plankton taxa into 85 different plankton groups, corresponding to the Major and Minor Groups of the COPEPOD PGC system.

Life Stage: Each taxonomic observation is split into one of five life stage categories. These categories are "adult", "larger sub-stages", "smaller sub-stages", "eggs", and "not specified". The main purpose for these groupings is to separate small life stage counts from counts of adults or larger life stages. While these smaller life stages may be present in larger mesh nets, they generally require smaller mesh sizes and additional counting effort to properly enumerate.

Taxonomic Hierarchy Level: This is a very important sub-category as counts for a single species will clearly have a different value range than a count (or sum) of an entire class or order (*i.e.,* composed of multiple species). Each taxa observation is split into seven taxonomic hierarchy categories: 1= species, 2= genus, 3= family, 4= order, 5 = class, 6 phylum, and 7=kingdom. "Super-", "infra-", and "sub-" extensions are ignored (*e.g.,* "suborder" = "superorder" = "order" = level 4). Regardless of the hierarchy category, any taxa description with a "sp." single-species indicator (*e.g., Calanus* sp., Copepod sp., Crustacean sp., Diatom sp.) is automatically assigned to level "1" (species level) as it is assumed to refer to a single, unidentified species from that group.

Geographic Region: Each data value is checked against other data from the same geographic region as well as all data from the entire world ("global"). The geographic regions used in *COPEPOD-2010* (Figure 4) are based on those used by the *World Ocean Database* for hydrographic and nutrient data regions (*e.g.,* Conkright *et al. 2002,* O'Brien *et al.* 1998), without any coastal/open-ocean designations.

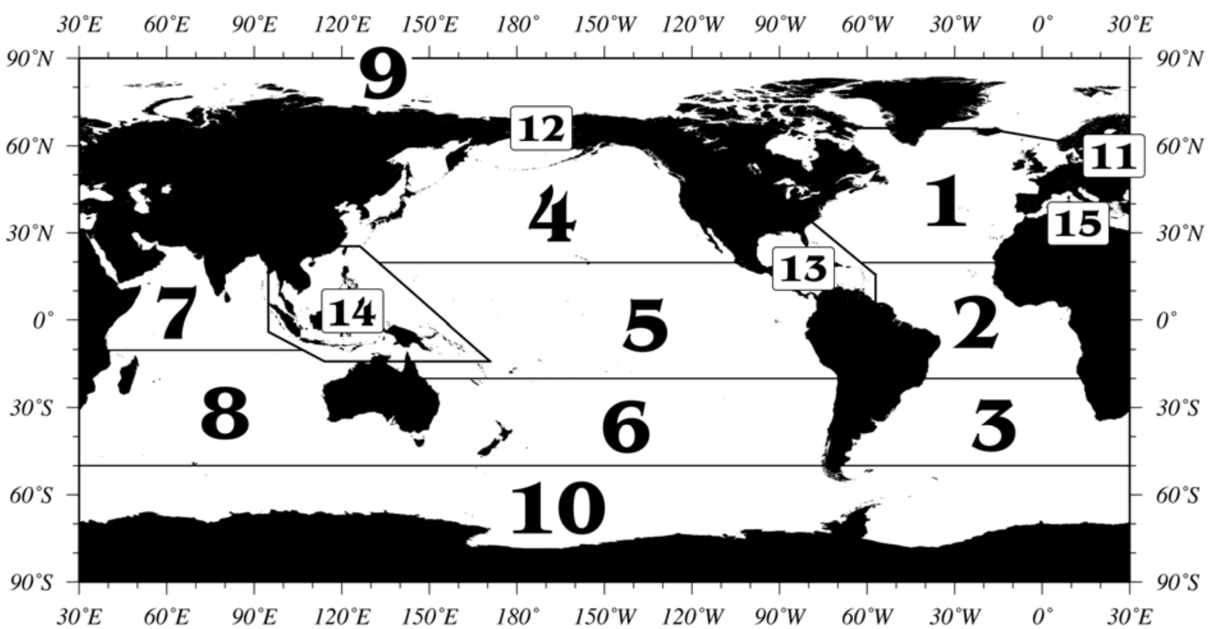


Figure 4: Map of geographic regions used by the *COPEPOD-2010* ranging system.

(1) North Atlantic, (2) Equatorial Atlantic, (3) South Atlantic, (4) North Pacific, (5) Equatorial Pacific, (6) South Pacific, (7) North Indian, (8) South Indian, (9) Arctic, (10) Antarctic, (11) Baltic, (12) Bering Sea, (13) Gulf of Mexico & Caribbean, (14) Indonesia, (15) Mediterranean.

Seasons: Each data value is checked against other data from the same "season" as well as all data from "any season or month" (annual). In *COPEPOD-2010*, and COPEPOD in general, seasons are defined as "winter" (December - February), "spring" (March - May), "summer" (June - August), and "autumn" (September - November).

Months: When data density allows, data values were also checked against other data from the same month. These checks are generally available in data rich areas such as the US coast lines, but not much of the data sparse open ocean regions (*e.g.,* South Pacific, Arctic region).

Time of Day: Each value is checked against other data from the same "day" or "night" category, where "night" is defined as one-hour after local sunset* until one-hour before local sunrise* and "day" is defined as one-hour after sunrise until one-hour before sunset. Local sunrise and sunset times were calculated using sample latitude, longitude, date, and time in the United States Naval Observatory Sunrise/Sunset Algorithm (*Almanac for Computers*, 1990). The one hour buffers surrounding sunrise and sunset were excluded to avoid samples taken during the day/night transition. Data from these transition periods, as well as any samples or data sets with missing times, were checked against a general "any time / all times" range (instead of the specific "day" or "night" range).

Net Mesh Size: Each observation is checked against other data from the same general net mesh size. *COPEPOD-2010* mesh size categories were based upon the frequency of employed mesh sizes in the main database and labeled per general target organisms. Raw mesh sizes were split into one of six size categories: Meso-zooplankton mesh sizes (200 µm, 300 µm, 500 µm), fine plankton mesh sizes (050 µm, 100 µm), and "bottle". Table 6 illustrates the mesh size categories and criteria used within *COPEPOD-2010*.

		Fine Mesh		Meso Mesh		
	"bottle"	"050 µm *25 - 76 µm*	"100 µm *94 - 125 µm*	"200 µm *150 - 253 µm*	"300 µm *270 - 417 µm*	"500 µm *470 - 570 µm*
Total Biomass	-	X	X	X	X	X
Bacterioplankton	X	*	*	*	*	*
Phytoplankton	X	X	X	*	*	*
Zooplankton	-	X	X	X	X	X
Ichthyoplankton	-	-	X	X	X	X

Table 6. Range checking mesh categories used within *COPEPOD-2010*.

* Phytoplankton from mesh sizes > 100 µm and non-bottle bacterioplankton are flagged as "gear bias" (*i.e.,* specimens periodically snagged on the net but generally smaller than the net mesh opening).

Combining Sub-Groups

In addition to splitting the data into sub-categories, the *COPEPOD-2010* ranging system combines individuals into “combined grouping sums” (CGS) to allow for additional ranging checks at higher group or taxa-hierarchy levels. For example: A new data collection has comprehensive individual copepod genus and species-level counts, but no other genus or species data are available for this specific region or season in the main database. The main database does, however, have thousands of “total copepod” counts in that same region and season. To see if the new data have reasonable value ranges, the COPEPOD ranging system can add together all PGC “copepod” counts within the new data to create a combined group sum (CGS) for “total copepods”. This CGS value can then be compared to the total copepod data, already present in the main database, to check for any gross value errors in the new data.

Future Sub-Categories

The COPEPOD ranging system is still in a developmental state. What is presented above is actively being applied to all *COPEPOD-2010* data, but additional sub-categories are being added and tested for incorporation in the near future. These improvements include:

SAMPLING DEPTH: The majority of the plankton data in COPEPOD are from vertical or oblique tows sampling from around 200 meters depth (or the bottom, if it is shallower) to the surface. Additional depth categories may be added to handle surface and/or depths to 500 meters.

IMPROVED GEOGRAPHIC REGIONS: In most regions, there is a visable difference between the productive near-shore (“coastal”) area and the typically oligotrophic open-ocean areas (*see* Figure 11 *in the Data Products section for an example*). In the same figure, we also see that regions such as the Equatorial Pacific have large differences between the western and eastern side of the basin. Incorporating “near shore vs. open ocean” sub-regions, or switching to a grid of standard geographic cells (*e.g.,* 5° x 5° latitude-longitude boxes*),* would better separate these productive and oligotrophic regions, allowing for tighter value ranging checks.

4.2 Value Ranging and Outlier Detection

For each ranging sub-category, simple statistics are run on all available COPEPOD data within that specific sub-category and used to create a “ranging set” for that specific sub-category. For example, season-based ranging statistics for “North Pacific total wet mass (mg/m^3) data sampled with a 300 µm mesh net” might look like Table 7.

PGC	Region	Season	Mesh	"n"	< 99.99%	< 99.9%	< 99%	Median	> 99%	> 99.9%	> 99.99%
-403	4	13	300	**2,894**	0.01	0.04	0.60	16.8	242	707	1,000
-403	4	14	300	**5,305**	0.01	0.25	0.84	58.7	545	979	1,200
-403	4	15	300	**12,202**	0.04	0.42	1.68	80.5	1,752	6,253	13,111
-403	4	16	300	**5,469**	0.01	0.13	0.42	37.7	298	730	1,769

Table 7: Example of ranging sets for Total Zooplankton Wet Mass data (*PGC = -403*) sampled with 300 µm mesh in the North Pacific (*Region = 4*). Shown are ranging sets for general winter (*Season = 13*), spring (*14*), summer (*15*), and autumn (*16*) time periods.

Each ranging set contains the minimum and maximum value limits for a data value to be considered within 99%, 99.9%, or 99.99% of all the other available data in the entire COPEPOD database. These ranging limits were selected so each category was ten-times larger than the previous category. If a data value falls outside of one of these ranging limits, that data value is flagged. The actual ranging flag assigned is dependent on a minimum “n” (the number of other data values available in the main database for that sub-category). At least 100 observation values are needed in the main database for a given sub-category for any ranging flag to be assigned by that sub-category. If 1,000 or more values are available, the tighter 99.9% ranging check can be assigned, and the even tighter 99.99% ranging flag can be assigned if 10,000 or more values are available. This “n”-based ranging and flagging is system summarized in Table 8.

	2010 Flag	“n” range *weaker*	 **stronger**
“Gear Bias”	-9	Applies to bacteria and phytoplankton (see Table MM).	
< 99.99% of all data	-7		“n” > 100,000
	-6	*“n” > 10,000*	
< 99.9% of all data	-5		“n” > 10,000
	-4	*“n” > 1,000*	
< 99% of all data	-3		“n” > 1,000
	-2	*“n” > 100*	
“Zero Value” indicator	-1	*Used for “zero values”.*	
“reasonable value”	**0**		
“n < 100” indicator	1	“n” < 100 (no ranging)	
> 99% of all data	2	*“n” > 100*	
	3		“n” > 1,000
> 99.9% of all data	4	*“n” > 1,000*	
	5		“n” > 10,000
> 99.99% of all data	6	*“n” > 10,000*	
	7		“n” > 100,000

Table 8: Table of *COPEPOD-2010* ranging flags and criteria.

The *COPEPOD-2010* ranging system assigns four ranging flags to each value. The first flag is a “global – annual” comparison flag, in which same sub-category data from all ocean regions and any season are used for the check. The second flag is a “regional – annual” comparison flag, in which same sub-category data from the same region (but any season) are used for the check. The third flag is a “regional – seasonal” comparison flag, in which same sub-category data from the same region and same season are used for the check. The fourth flag is a “regional – monthly” comparison flag, in which same sub-category data from the same region and same month are used for the check. (Again, if less than 100 values for this region and month are available, flagging is not assigned for that region-month.)

General Comments about flag results:

- The purpose of the ranging flags is to allow for a quick and general indication of how a data value compares to similar data values found in the entire COPEPOD database. The data are flagged, but not removed, allowing the data user to use or ignore the flags by their own choice.

- The higher the flag value (*or the lower, in the case of negative flags*), the more likely it is that the data are non-representative (anomalous). For example, a flag value of "7" means that this specific data value is larger than 99.99% of at least 100,000 other existing data values for the exact same plankton measurement. While it could still be a legitimate value, it is probably worth investigating before using it.

- It is possible for value to have different ranging flags assigned for each of the three flags. For example, a higher spring value may be flagged as a ">99%" global value, but be perfectly reasonable within its regional and seasonal sub-categories. Likewise, a specific basin or season may have limited data, so a weaker or "n <100" flag may be assigned for the regional or seasonal check, but a stronger n-based flag could be assigned by the global check.

- Having a few values outside of the 99% limit is usually okay, especially if the bulk of the other data falls within those 99% limits. If large portions of the data flagged as >99%, or if multiple >99.9% or >99.99% flags are present, the data should be investigated.

4.3 Value Frequency and Range Visualization

The COPEPOD ranging system generates standard results tables for each data sub-category. Each table reports all plankton groups present in the data collection, listing the number of observations and percentage of those observations flagged for each flag type, similar to Table 9.

PGC	Region	Season	Mesh	*"n"*	**<99.99%**	<99.9%	*<99.0%*	**"okay"**	*>99.0%*	>99.9%	**>99.99%**
-403	**WORLD**	***Annual***	300	485	.	0.60%	.	99.00%	*0.40%*	.	.
-403	*N.Pacific*	***Annual***	300	485	.	0.60%	.	99.00%	*0.40%*	.	.
-403	*N.Pacific*	*Winter*	300	47	.	.	.	**100.00%**	.	.	.
-403	*N.Pacific*	*Spring*	300	90	.	3.30%	.	95.60%	*1.10%*	.	.
-403	*N.Pacific*	*Summer*	300	309	.	.	*0.60%*	99.00%	*0.30%*	.	.
-403	*N.Pacific*	*Autumn*	300	39	.	.	.	97.40%	*2.60%*	.	.
4320000	**WORLD**	***Annual***	300	117	.	.	.	**100.00%**	.	.	.
4320000	*N.Pacific*	***Annual***	300	117	.	.	.	**100.00%**	.	.	.
4320000	*N.Pacific*	*Winter*	300	0	.	.	.	.	.	.	.
4320000	*N.Pacific*	*Spring*	300	0	.	.	.	.	.	.	.
4320000	*N.Pacific*	*Summer*	300	117	.	.	.	**100.00%**	.	.	.
4320000	*N.Pacific*	*Autumn*	300	0	.	.	.	.	.	.	.

Table 9: Example of COPEPOD ranging results for *Total Wet Mass* biomass data (*PGC = -403*) and *Chaetognath* abundance data (*PGC= 4320000*) sampled in the North Pacific with ~300 µm mesh nets.

The results tables can be annotated for quick, "at-a-glance" reviewing. In Table 9 the flagged-data columns are color-coded with "okay" values shown in green and flagged values are shown in gray or red. One can quickly scan through the entire results file, looking for instances of red to investigate. In this example, the Chaetognath data are all "okay" (green), all values fall within the 99% range, but some of the Total Wet Mass data have really low values (red, "<99.9%" values). There were 485 total wet mass tows in this example, most of them (309) made in the summer. While some data were flagged as >99% or <99% (gray), they are only a small portion of then entire data set and seem reasonable. The red flags of "< 99.9%", however, should be examined (*see below for further discussion*).

While Table 9 is useful for quickly reviewing a data set, plotting and viewing the same results can improve the ability to quickly review the data. Figure 5 is a visualization of the results shown in Table 9.

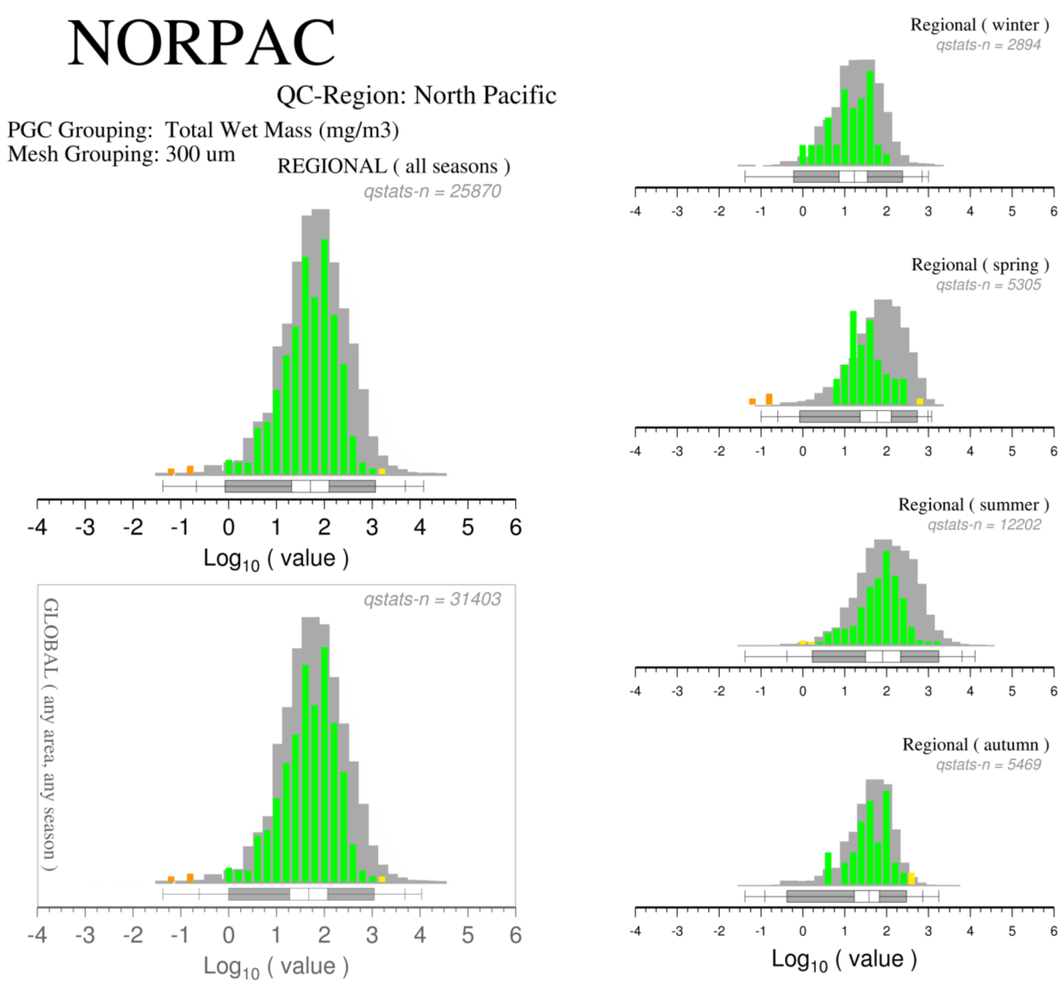


Figure 5: Ranging visualization for North Pacific *Total Wet Mass* data from the NORPAC project. The green/yellow/orange bars show NORPAC value distributions. The gray bars show value distributions for all corresponding wet mass data present in the entire COPEPOD database ("*qstats-n*" indicates the number of these values).

An explanation of Figure 5 (shown on previous page):

Bottom left sub-figure: Comparison of the NORPAC data to over 30,000 other “300 µm mesh wet mass values” (from any region, any season) present in the COPEPOD database.

Top left sub-figure: Comparison of the NORPAC data to over 25,000 other “300 µm mesh North Pacific wet mass values” (from any season) present in the COPEPOD database.

***Right column of four sub-figures*:** Comparison of the NORPAC data to other North Pacific ~300 µm mesh wet mass values from the same season.

In each figure, the box-n-whiskers portion immediately below the histograms indicates the ranging limits for 99% (gray outer box), with whiskers indicating the 99.9% (inner tick) and 99.99% (outer tick) limits. The white central box indicates the range of 50% of all data, the inner-most tick indicating the median of all data values.

The histograms in Figure 5 show the frequency of $\log_{10}$ total wet mass values. The two orange bars (seen in the global-annual, regional-annual, and regional-spring plots) are roughly 1000 times ($\log_{10}$ = “3”) smaller than the bulk of the other data values. (These are the same values indicated with the red “<99.9%” in Table 9.) Upon checking the original data source for these data, one would find an asterisk (*) next to the low values indicating gear problems for the three tows making up these points.

In general, the isolated values caused by a tow through a bloom or random equipment failure will show up as individual spikes (as seen in the orange bars of Figure 5). Systematic errors, where an entire group of values is too high or too low due to mistranslated units or a decimal error, will typically show up as a shift in some or all of the values in the histogram. For example, the shifted peak of red values in Figure 6 below turn out to be from the same cruise of a multi-cruise data set. In the documentation for that cruise was a note that all data from that specific cruise were “per 100 m^3”, a change from all previous cruises’ units in the same data set. By missing this one-time adjustment to the units, the data were accidentally loaded at 100 times their correct value. When plotted on the $\log_{10}$ scale, this 100x error appears as a $\log_{10}$ shift of “2” as seen in the red cluster in Figure 6.

The error mentioned above (and shown in Figure 6) was first discovered using the *COPEPOD-2007* quality control method. Prior to that it had remained undetected by the broad, range-based value checks used in the WOD-2001, WOD-2005, and COPEPOD-2005 database versions.

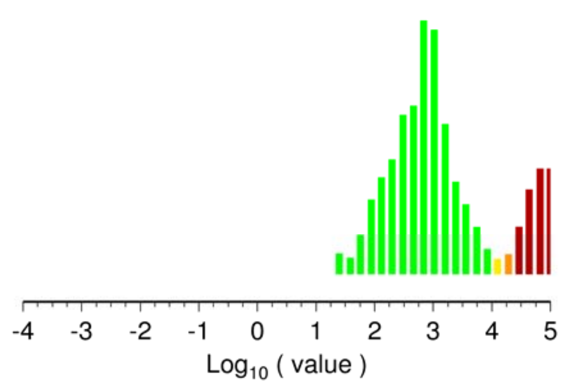


Figure 6: Example of a visual shift caused by systematic errors during data translation.

5. DATA PRODUCTS

In addition to offering hundreds of individual data collections, COPEPOD also offers regional and taxonomic compilations and data products which allow a user to work with the COPEPOD plankton data at a variety of geospatial scales (*e.g.*, local, regional, or global) and aggregate processing levels (*e.g.*, raw individual, species compilations, gridded aggregates and analyzed mean values).

5.1 Online Database and Interface

The online COPEPOD database (http://www.st.nmfs.noaa.gov/plankton) is the entry point for access to all of the COPEPOD data collections, compilations, and data products. Data collections can be easily searched for by their associated geographic regions, projects, research vessels or cruise, or even credited investigators. Data are available in both an abbreviated “spreadsheet-friendly” format and in a comprehensive “all-information” metadata and full content format.

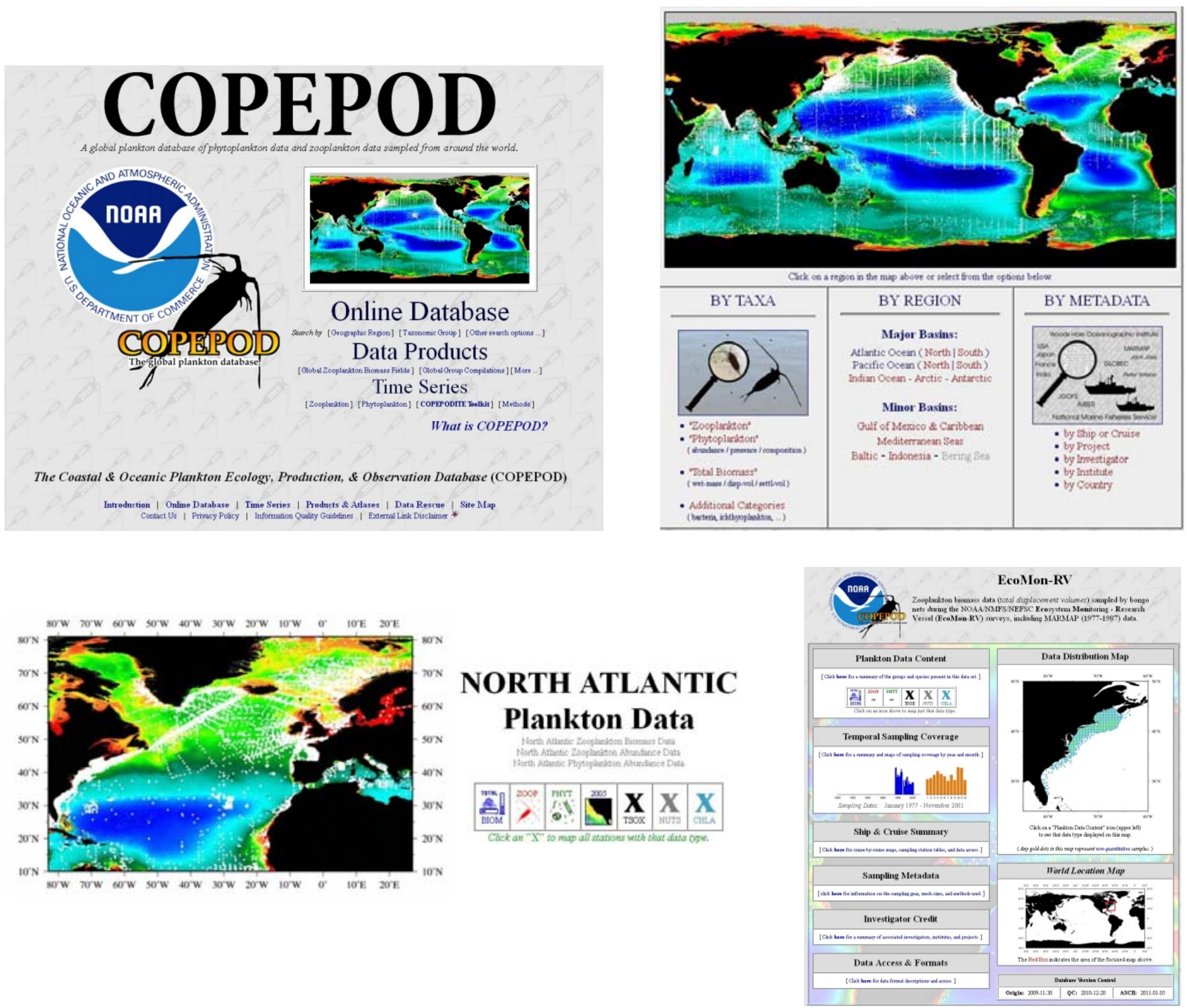


Figure 7: The COPEPOD online interface offers a variety of searching, access, and visualization options.

The COPEPOD approach to data management is to focus on individual data sets, highlighting each entity with a detailed summary of the exact data content, sampling methods, and investigators associated with those data. Each individual COPEPOD data set, called a "collection", includes an inter-linked, multi-page, html-based graphics and text content summary (Figures 8-9).

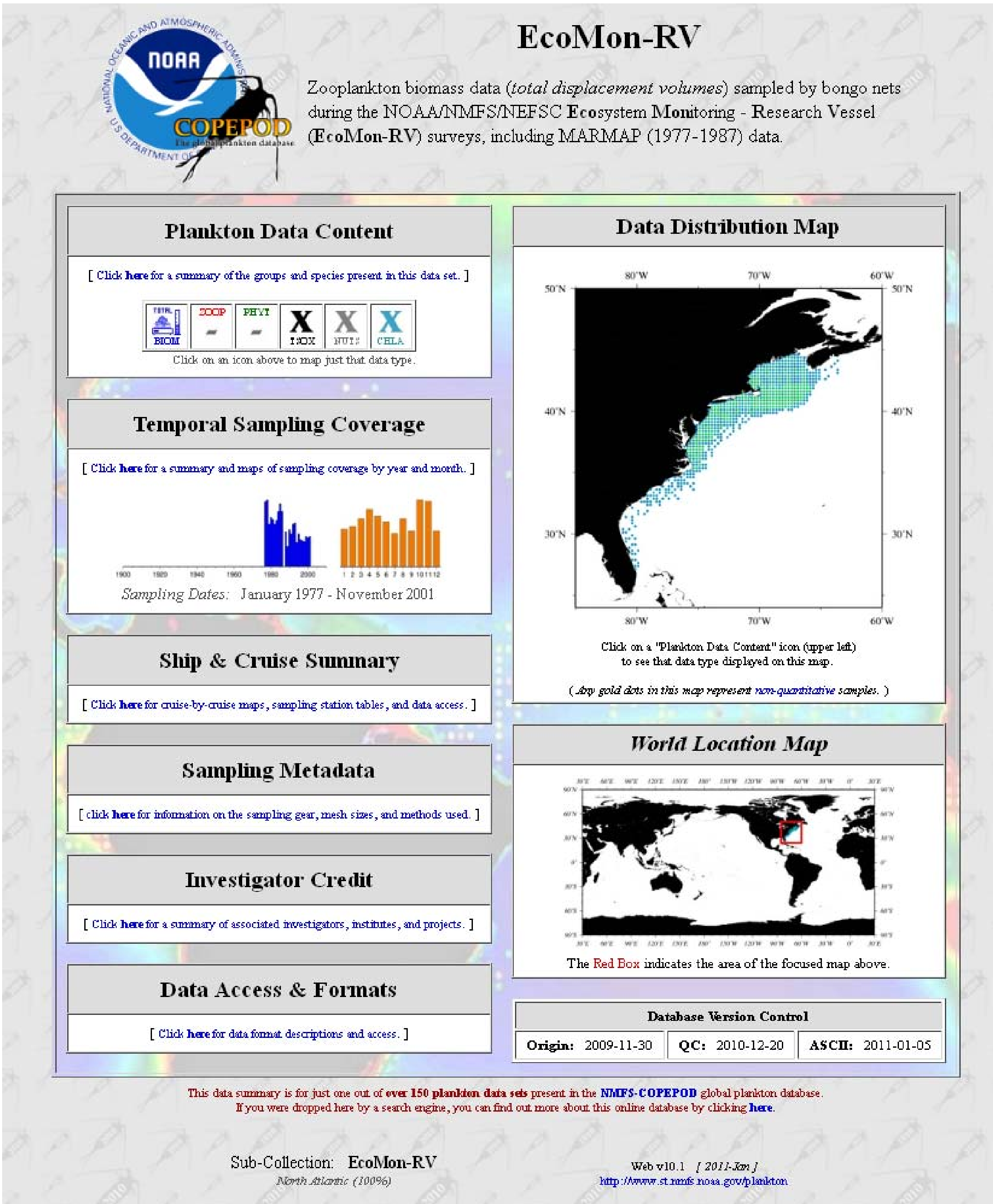


Figure 8: Example of a typical data collection summary main page, featuring the Research Vessel (RV) component of the NOAA / NMFS / NEFSC Ecosystem Monitoring (EcoMon) program.

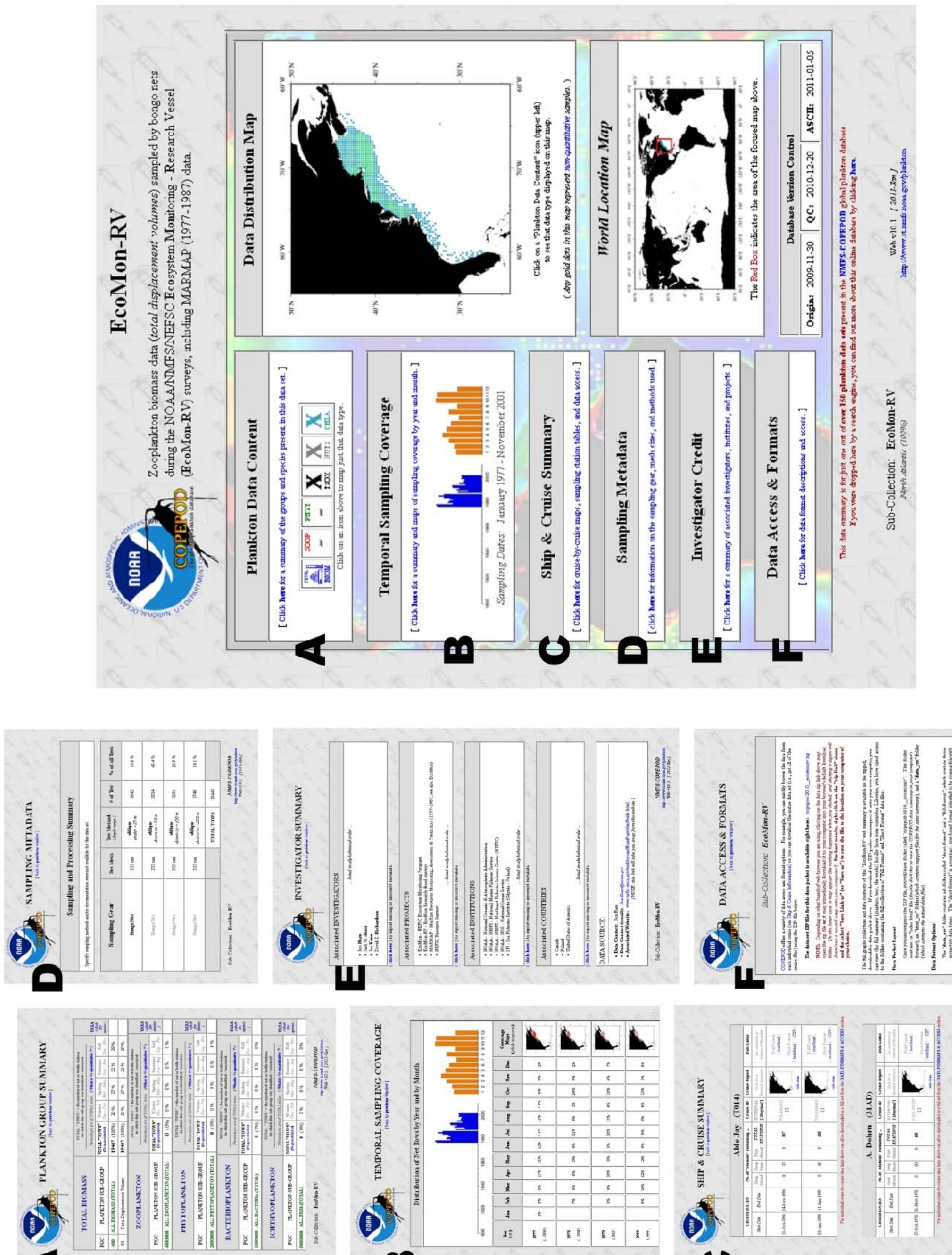


Figure 9: Examples of linked, sub-page summaries available within each COPEPOD data collection.

Each Collection Summary contains the following information and sub-pages:

A The six icons in the **Plankton Data Content** section show what general data types are available (*e.g.,* zooplankton, phytoplankton, hydrographic data, nutrient data). Clicking on one of these icons will plot that group's data coverage in the main ***Data Distribution Map*** (upper right of the main summary page). The blue web link in the Plankton Data Content section will take you to the ***"Plankton Group Summary"*** page. This sub-page summarizes all of the plankton groups and data available in the data collection. Additional links in these tables show listings of all species present in the collection.

B The **Temporal Sampling Coverage** section features a "data-by-years" histogram on the left and a "data-by-months" histogram on the right. The blue web link in this section will take you to the ***"Temporal Sampling Coverage"*** page. This sub-page features a table showing all data in the collection by year and individual month. Maps of data coverage for each individual year are also presented in this sub-page.

C The **Ship & Cruise Summary** section features a blue web link which takes you to the detailed ***"Ship & Cruise Summary"*** page. This sub-page features a table for each ship and lists each cruise (above that vessel) present in the data collection. Each row in these tables lists the data types present in the cruise, features a data distribution map for each cruise, and ***provides a link to the individual COPEPOD cruise-based data files available for that specific cruise***. This last feature lets you quickly review the actual data to see if it has the specific variables or species you may desire. (You can also download the entire data set through the link in section "F" below.)

D The **Sampling Metadata** section features a blue web link which takes you to the ***"Sampling Metadata"*** page. This sub-pages summarizes the sampling methods (in a descriptive paragraph form) as well as quantitative listings of the sampling gear, net mesh sizes, and towing methods applied within that data collection.

E The **Investigator Credit** section features a blue web link to the ***"Investigator Summary"*** page. This sub-page summarizes all known investigators, institutions, countries, projects, and data sources associated with the data collection. When available, this section provides contact information (email, institutional web page) and citation information for the data collection.

F The **Data Access & Formats** section features a blue web link to the ***"Data Access & Formats"*** page. This sub-page provides documents and provides information on the available data formats and data structures. A link to a ZIP file in this section lets you download the entire data collection (including all of the data as well as a local copy of the web summary, metadata, and graphics).

5.2 Online Interactive Atlas

COPEPOD-2010 features an electronic atlas of data distribution maps and species lists for each of the major Plankton Groups shown in Table 3. Within each map, a graphical dot indicates the presence of at least one tow (or bottle sample) in the main database which has observations for that plankton group. A summary table lists the name of all species or measurement types present in the database, listed in order of frequency. Finally, users can download a global compilation containing all available data for that specific plankton group.

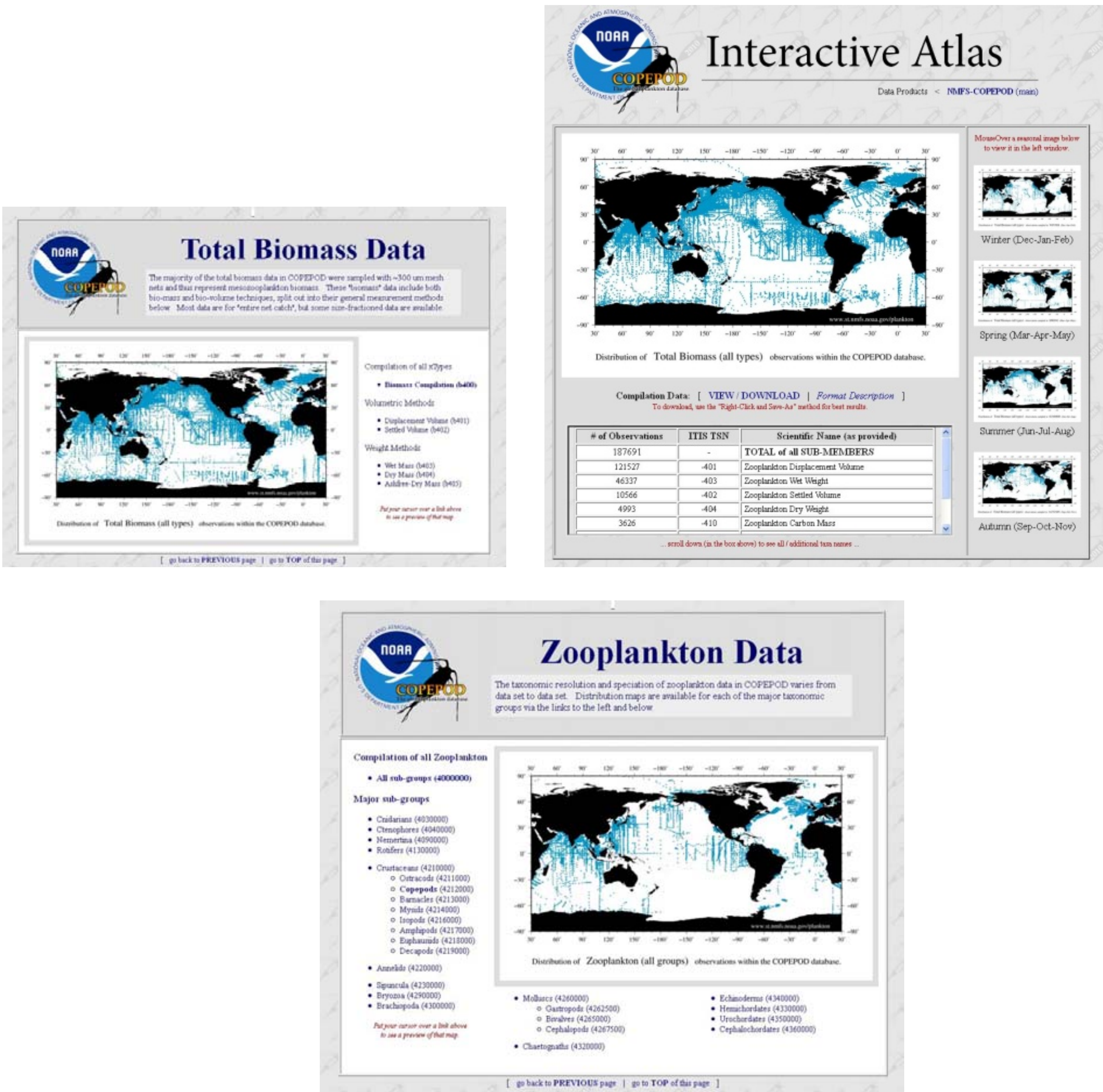


Figure 10: Examples of the data content tables and plots found in the online Atlas section.

5.3 Global Zooplankton Biomass Fields

The global zooplankton biomass fields of *COPEPOD-2005* have been recreated to include new data from the Arctic, Antarctic, and South Pacific regions. These new fields were created using the same depth ranging, mesh adjustments, and carbon calculation equations used for *COPEPOD-2005* (O'Brien 2005), but now use the *COPEPOD-2010* ranging system to detect, investigate, and/or exclude any data values flagged as >99.9% or <99.9% within their respective region and season. This additional step was found to help exclude very large values (typically caused by large phytoplankton blooms clogging the zooplankton net) and very low values (typically caused by gear failures).

The online *COPEPOD-2010* biomass fields include annual and seasonal mean carbon mass (mg-C/m^3), total wet mass (mg/m^3), total displacement volume (ml/m^3), total settled volume (ml/m^3), and total dry mass (mg/m^3).

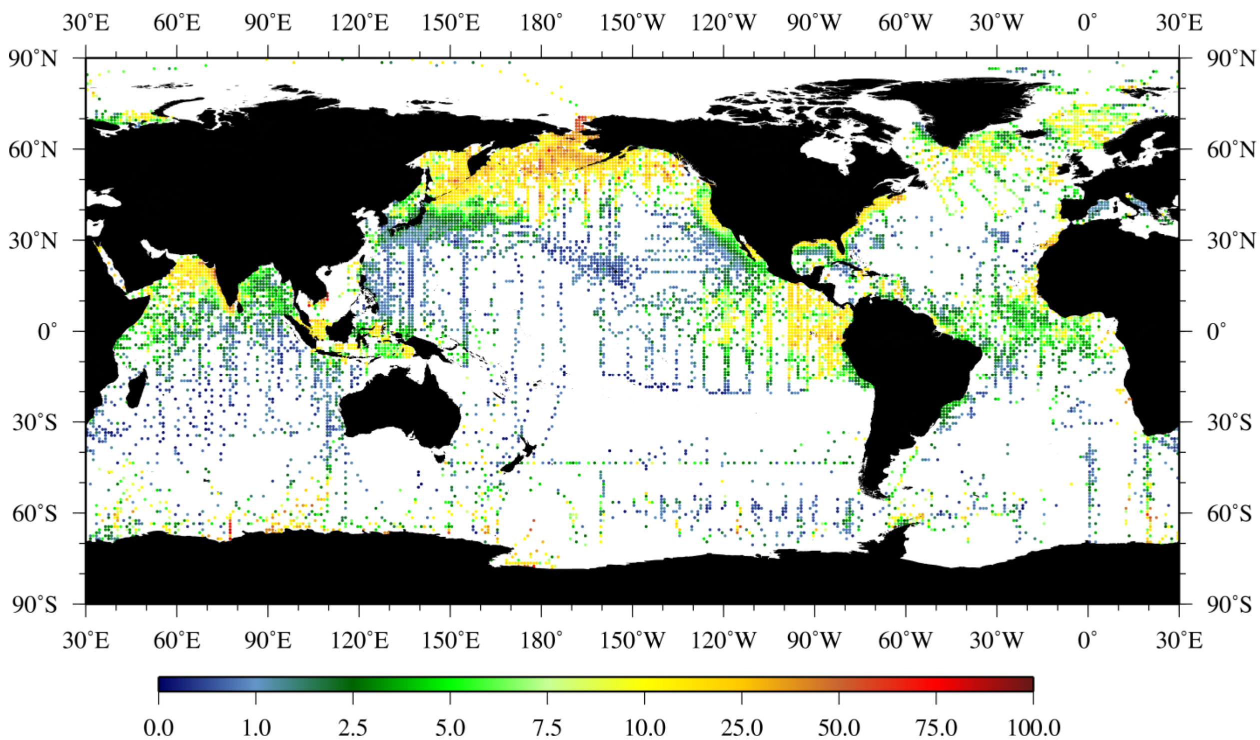


Figure 11: Annual mean zooplankton carbon mass (mg-C/m^3) fields from *COPEPOD-2010.*

REFERENCES

Conkright, M.E., H.E. Garcia, T.D. O'Brien, R.A. Locarnini, T.P. Boyer, C. Stephens, J.I. Antonov, 2002. World Ocean Atlas 2001, Volume 4: Nutrients. NOAA Atlas NESDIS 52, U.S. Government Printing Office, Wash., D.C., 392 pp.

Margulis, L. and K.V. Schwartz 1998. Five Kingdoms: An Illustrated Guide to the Phyla of Life on Earth. W.H. Freeman & Company (New York), 520 pp.

O’Brien, T.D. 2007. COPEPOD: The Global Plankton Database. A review of the 2007 database contents and new quality control methodology. U.S. Dep. Commerce, NOAA Tech. Memo. NMFS-F/ST-34, 28 pp.

O’Brien, T.D. 2005. COPEPOD: A Global Plankton Database. A review of the 2005 database contents and creation of new global zooplankton biomass fields. U.S. Dep. Commerce, NOAA Tech. Memo. NMFS-F/SPO-73, 136 pp.

O'Brien, T.D., M. E. Conkright, T. P. Boyer, J. I. Antonov, O. K. Baranova, H. E. Garcia, R. Gelfeld, D. Johnson, R. A. Locarnini, P. P. Murphy, I. Smolyar, C. Stephens 2002. World Ocean Database 2001, Volume 7: Temporal Distribution of Chlorophyll and Plankton Data. NOAA Atlas NESDIS 48, U.S. Gov. Printing Office, Wash., D.C., 219 pp.

O' Brien, T.D., M.E. Conkright, T.P. Boyer, D. Johnson, C. Stephens, R. Gelfeld 1998. World Ocean Database 1998 Volume 7: Temporal Distribution of Oxygen, pH, and Alkalinity Profiles. NOAA Atlas NESDIS 24, U.S. Gov. Printing Office, Wash., D.C., 235 pp.

Listing of Investigators associated with the "*COPEPOD-2010*" data content*.

[A] - T. ABE - Quaraisin ADNAN - A.P. ALEKSEEV - Arshad ALI - Moazzarn ALI – P.J. ALLEN - Usman ALWI - M. AMADEO - Shahid AMJAD – J. DUFFY-ANDERSON - W.W. ANDERSON - Oscar ANGELES-REYES - Michel ANGOT - P. AOMORI - E. AOYAMA - Dharma ARIEF - O. H. ARINARDI - E. ASADA - Carin J. ASHJIAN - H. AUEL - Farooq AZAM - **[B]** - V. BAINBRIDGE - E.P. BALDINA - Richard T. BARBER - Richard A. BARKLEY - Clifford A. BARNES - Sonia BATTEN - Timothy R. BAUMGARTNER-McBRYDE – Allan BÉ - J. BEAUDOUIN - Robert L. BENWAY - T. BERNSHTEIN - Robert BIDIGARE - O.K. BILJEVA - George BOEHLERT - V. G. BOGOROV - Ph. BOURRET - E. BRATBERG - K.A. BRODSKII - Robert P. BROWN - E. BUSKEY - **[C]** - C. CABOCHE - Lisa CAMPBELL - H. CARADANT - David CARON - Louis CODISPOTI - D.V.P. CONWAY - R.T. COONEY - J. CORLETT - Charles S. COX - Maria M. CRIALES - Henryk CZKIETA - **[D]** - Michael DAGG - Donald S. DAY - K.M. DERIJUGIN - DJUMHAR - Maxwell S. DOTY - Hugh DUCKLOW - Janet DUFFY-ANDERSON - **[E]** - S. EBARA - Sayed Z. EL-SAYED - Bjornar ELLERTSEN - T. Saunders ENGLISH - **[F]** – Frank FERRARI – J. FLINKMAN - P. FOXTON - Mitsuo FUKUCHI - Y. FUKUDA - P. FUKUOKA - K. FURUHASHI - **[G]** - David GARRISON - Anne GAUZENS - J.W. GEHRINGER - R.S. GLOVER - G.P. GORBUNOV - Michel GOSSELIN - Julien R. GOULET - Marcia GOWING - E.H. GRAINGER - John R. GREEN - P.R.S. GREEN - Brian GRIFFITHS - J.R. GRINDLEY - Jean-Alfred GUEREDRAT - **[H]** - W. HAGEN - I. HALLGRIMSSON - H. HAMADA - Y. HAMADA - M. HANDA - Vagn HANSEN - H. HANSSEN – Jon HARE - P.M. HARGREAVES - H. HATTORI - Ken-Ichi HAYASHI - Erica HEAD - Thomas S. HIDA - Harry HIGGINS - Hans-Jurgen HIRCHE - Robert W. HOLMES - Rita HORNER - Takao HOSHIAI - L. HUTCHINGS - P. HYOUGO - **[I]** - A. G. ILAHUDE - Gennady V. ILYIN - M. IMAI - Yoshio INO - M. ISHII - Shingo ISHIKAWA - **[J]** - Luis Clemente JIMENEZ-PEREZ - P.L. JOHANSEN - Martin W. JOHNSON - Jack W. JOSSI - **[K]** - H. KAMAZAKI - M.M. KAMSHILOV - Hiroshi KANDA - Joe KANE - K. KAROHJI - Akiyoshi KATAOKA - Gerhard KATTNER - M. KAWACHI - S. KAWAE - Kouichi KAWAGUCHI - K. KAWASHIMA - KENDAL - V.L. KHMIZNIKOVA - I.A. KISELEV - Wojciech KITTEL - Anthony H. KNAPP - A.N. KOLESNIKOV - K. KONDOU - T. KONNO - A. KOTTHAUS - N. KUBO - T. KUBODERA - T. KURAMOCHI - K. KURODA - Leonid KUZNETSOV - **[L]** - Mike LANDRY - Peter V.Z. LANE - S.B. LANE - R.J. LARA - Viktor V. LARIONOV - Bertha E. LAVANIEGOS - Frederik LEATEMIA - Jacques LEBOURHIS – Maiju LEHTINIEMI - Bill LI - Alan LONGHURST - A.N. LUSHOV - Waldo K. LYON - **[M]** - David L. MACKAS - Enrique F. MADELLI - Pavel R. MAKAREVICH - Isak MALIK - T. MANABE - T.O. MANLEY - K.H. MANN - Ogur MARPAUNG - John MARRA - Osamu MATSUDA - Y. MATSUDA - M. MATSUZAKI - T. MAYAMA - John A. McGOWAN - Skip McKINNELL - Edison MEILALA – Webjorn MELLE - T. MINODA - Y. MIYAMOTO - MOESOLL - E. MORIYAMA - Robert A. MORRIS - V.K. MORYAKOVA - Shigeru MOTODA - Kojiro MOTOMURA - Muswerry MUCHTAR - Nicolai MUMM - Erik MUXAGATA - **[N]** - Yasushi NAKAJIMA - Eugene L. NAKAMURA - T. NAKAMURA - M. NANDA - V.A. NESMELOVA - R.N. NISHIMOTO - K. NOMURA - Anugerah NONTJI - **[O]** – Todd D. O'BRIEN - Kazuko ODATE - T. ODATE - Mikiho OGURI - M. OHNO - Hideaki OHTSUKA - Y. OHYAMA - Robert OLSON – Lana ONG - K. ONISHI - Jay E. O'REILLY - Y. OSAKA - **[P]** - Taisoo PARK - La PAY - Chairil PELHOUPESSY – Bill PETERSON - PHIFE - Sergey PIONTKOVSKI - Ewa PRESLER - I. Yu. PRUSOVA - V.P. PTOKHOV - **[Q]** - D. QUEIROZ - **[R]** - T.S.S. RAO - R. REPELIN – D.E. RICHARDSON - Christoph RICHTER - J. RIVATON - G.A. ROBINSON - Stephen J. ROMAINE - Mike ROMAN - ROUKHIY - Maria I. ROUKHIYAINEN - Jury RUDJAKOV - RUSMINTARD - **[S]** - H. SAKAI - Jennifer SALTZMAN - Daniel SAPULETE - Hiroshi SASAKI - N. SATO - Hiroo SATOH - M.S. SAVICH - Vladimir M. SAVINOV - R. SEIKAI - R.A. SELVAKUMAR - Kenneth SHERMAN - Alexandra A. SHMELEVA - Daniel E. SMITH - David SMITH - Sharon SMITH - Walker O. SMITH - Walter O. SMITH - Eddy SOFYAN - A.A. SOLOVJOVA - Grieg STEWARD - SUDIRDJ - M. SUGIMOTO - Yukio SUGIMURA - SUMARS - Vijayaraghavan SUMITRA - SUNARY - SUPRAPT - SUSEN - A. B. SUTOMO - SUYANT – Toru SUZUKI - [**T]** - Eiji TAKAHASHI - Kazutaka TAKAHASHI - I. Yu. TAMOIKIN - S. TAMURA - Akira TANIGUCHI - Atsushi TANIMURA - Persadaan TARIGAN - David TERHELL - Delphine THIBAULT - T.G. THOMPSON - Richard THOMSON - H. THUTHUI - U. TILMANN - Sergey F. TIMOFEEV - R. TSUDA - Arao TSURUTA - D.Th. TUHUMENA - **[U]** - Syunshiro UENO - Juanita URBAN-RICH - P.I. USATSCHEV - **[V]** - M. VANNUCCI - M.A. VIRKETIS - **[W]** - Brenda J. WADDELL - WAHY - K. D. WALDRON - Kenneth D. WALDRON - Kentaro WATANABE - Y. WATANUKI - Benny WATTIMENA - G. WEISSBERG - David W. WELCH - Pat WHEELER – Peter WIEBE - Karen WISHNER - Zbigniew WITEK - Sam WOUTHUYZEN - **[Y]** - Ken-ichi YAMAMOTO - H. YAMAZAKI - J. YANO - M. YASUDA - G. YOKOE - M. YOSHIDA - Salili Ali YUSUF - **[Z]** - M.M. ZABELINA - Angelina A. ZELIKMAN - Maria ZMIJEWSKA - Aleksey ZUYEV

* This list is based names provided with the respective data sets. This listing is incomplete as not all data contains full information on the associated investigators. If a name is missing from this listing, please contact: Todd.Obrien@noaa.gov

NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION

NOAA

U.S. DEPARTMENT OF COMMERCE

COPEPOD

The global plankton database!